

常见的基于图结构的知识图谱推理算法。

2. 常见算法简介

典型的基于图结构的推理方法有 PRA (Path Ranking Algorithm)^[10]利用了实体节点之间的路径当作特征从而进行链接预测推理。

(1) 基于知识图谱路径特征的 PRA 算法。PRA 处理的推理问题是关系推理，其中包含了两个任务，一个是给定关系 r 和头实体 h 预测可能的尾实体 t 是什么，即在给定 h, r 的情况下，预测哪个三元组 (h, r, t) 成立的可能性比较大，叫作尾实体链接预测；另一个是在给定 r, t 的情况下，预测可能的头实体 h 是什么，叫作头实体链接预测。

PRA 针对的知识图谱主要是自底向上自动化构建的含有较多噪声的图谱，例如 NELL，并将关系推理的问题形式化为一个排序问题，对每个关系的头实体预测和尾实体预测都单独训练一条排序模型。PRA 将存在于知识图谱中的路径当作特征，并通过图上的计算对每个路径赋予相应的特征值，然后利用这些特征学习一个逻辑斯蒂回归分类器完成关系推理。在 PRA 中，每一个路径可以当作对当前关系判断的一个专家，不同的路径从不同的角度说明了当前关系的存在与否。

在 PRA 中，利用随机游走的路径排序算法首先需要生成一些路径特征，一个路径 P 是由一系列关系组成的，即：

$$P = T_0 \xrightarrow{r_1} T_1 \xrightarrow{r_2} \dots \xrightarrow{r_{n-1}} T_{n-1} \xrightarrow{r_n} T_n.$$

式中， T_n 为关系 r_n 的作用域 (range)以及关系 r_{n-1} 的值域 (domain)，即 $T_n = \text{range}(r_n) = \text{domain}(r_{n-1})$ ，关系的值域和作用域通常指的是实体的类型。基于路径的随机游走定义了一个关系路径的分布，并得到每条路径的特征值 $s_{h,P(t)}$ ， $s_{h,P(t)}$ 可以理解为沿着路径 P 从 h 开始能够到达 t 的概率。具体操作为，在随机游走的初始阶段， $s_{h,P(e)}$ 初始化为1，如果 $e = s$ ，否则初始化为0。在随机游走的过程中， $s_{h,P(e)}$ 的更新原则如下：

$$s_{h,P(e)} = \sum_{e' \in \text{range}(P_i)} s_{h,P'(e')}(e') \cdot P(e|e'; r_i).$$

式中， $P(e|e'; r_i) = \frac{r_i(e', e)}{|r_i(e', \cdot)|}$ 表示从节点 e' 出发沿着关系 r_i 通过一步的游走能够到达节点 e 的概率。对于关系 r ，在通过随机游走得到一系列路径特征 $P_r = \{P_1, \dots, P_n\}$ 之后，PRA 利用这些路径特征为关系 r 训练一个线性的预测实体排序模型，其中关系 r 下的每个训练样本，